Optimal Privacy-aware Path in Hippocratic Databases *

Min Li[1], Xiaoxun Sun[1], Hua Wang[1], and Yanchun Zhang[2]

[1] Department of Mathematics & Computing
University of Southern Queensland, Australia
Email: {limin,sunx,wang}@usq.edu.au
[2] School of Computer Science & Mathematics
Victoria University, Australia
Email: yzhang@csm.vu.edu.au

Abstract. Privacy becomes a major concern for both customers and enterprises in today's corporate marketing strategies, many research efforts have been put into developing new privacy-aware technologies. Among them, Hippocratic databases are one of the important mechanisms to guarantee the respect of privacy principles in data management, which adopt purpose as a central concept associated with each piece of data stored in the databases. The proposed mechanism provides basic principles for future database systems protecting privacy of data as a founding tenet. However, Hippocratic databases do not allow to distinguish which particular method is used for fulfilling a purpose. Especially, the issues like purpose hierarchies, task delegations and minimal privacy cost are missing from the proposed mechanism.
In this paper, we extend these mechanisms in order to support inter-organizational business processes in Hippocratic databases. A comprehensive approach for negotiation of personal information between customers and enterprises based on user preferences is developed when enterprises offer their clients a number of ways to fulfill a service. We organize purposes into purpose directed graphs through AND/OR decomposition, which supports task delegations and distributed authorizations. Specially, customers have controls of deciding how to get a service fulfilled on the basis of their personal feeling of trust for any service customization. Quantitative analysis is performed to characterize privacy penalties dealing with privacy cost and customer's trust. Finally, efficient algorithms are given to guarantee the minimal privacy cost and maximal customer's trust involved in a business process.

1 Introduction

With the widespread use of information technology in all walks of life, personal information is being collected, stored and used in various information systems. Achieving privacy preservation has become a major concern. Issues related to

* This research is funded by an ARC Discovery Project DP0663414.

privacy have been widely investigated and several privacy protecting techniques have been developed. To our best knowledge, the most well known effort is the W3C's Platform for Privacy Preference (P3P) [10]. P3P allows websites to express their privacy policy in a machine readable format so that using a software agent, consumers can easily compare the published privacy policies against their privacy preferences. While P3P provides a mechanism for ensuring that users can be informed about privacy policies before they release personal information, some other approaches are proposed [4, 8, 16, 17, 22], where the notion of *purpose* plays an important role in order to capture the intended usage of information.

As enterprises collect and maintain increasing amounts of personal data, individuals are exposed to greater risks of privacy breaches and identity theft, many enterprises and organizations are deeply concerned about privacy issues as well. Many companies, such as IBM and the Royal Bank Financial Group, use privacy as a brand differentiator [3]. By demonstrating good privacy practices, lots of business try to build solid trust to customers, thereby attracting more customers [20, 5, 6, 12]. Together with the notion of *purpose*, current privacy legislations also define the privacy principles that an information system has to meet in order to guarantee customer's privacy [11, 1, 2, 21]. Mechanism for negotiation is presented by Tumer et al. [21]. Enterprises specify which information is mandatory for achieving a service and which is optional, while customers specify the type of access for each part of their personal information.

On the basis of the solution for the exchange between enterprises and customers, Hippocratic databases enforced fine-grained disclosure policies to an architecture at the data level [1]. In the proposed architecture, enterprises declared the purpose for which the data are collected, who can receive them, the length of time the data can be retained, and the authorized users who can access them. Hippocratic databases also created a privacy authorization table shared by all customers, but it does not allow to distinguish which particular method is used for fulfilling a service. Moreover, enterprises are able to provide their services in different ways, and each different method may require different data. For example, notification can be done by email or by mobile phone or by fax. Depending on the different kinds of methods, customers should provide different personal information. Asking for all personal information for different service methods as compulsory would clearly violate the principle of minimal disclosure.

On the server side, a single enterprise usually could not complete all procedures of a service by itself, rather a set of collaborating organizations participating in the service. Enterprises might need to decompose a generic purpose into more specific sub-purposes since they are not completely able to fulfill it by themselves, and so they may delegate the fulfillment of sub-purposes to third parties. It is up to customers to decide on a strategy of how to get a service fulfilled on the basis of their personal feeling of trust for different service components. A question that many customers have when interacting with a web server, with an application, or with an information source is "Can I trust this entity?". Different customizations may require different data for which considerations may vary; there might be different trust levels on different partners

(sub-contractors). The choice of service customization has significant impact on the privacy of individual customers.

In this paper, we present an approach to automatically derive the optimal way of authorizations needed to achieve a service from enterprise privacy policies. In particular, we organize purposes into purpose directed graphs through AND/OR decompositions, which support the delegation of tasks and authorizations when a host of partners participating in the business service provides different ways to achieve the same service. Further, we allow customers to express their trust preferences associated with each partner of the business process. Thus, a weight combining privacy cost and customer trust is given on each arc of the graph in the form of privacy penalties, and the process for fulfilling a purpose can be customized at runtime and guarantees minimal privacy cost and maximal customer trust because it was selected with criterion of the optimal privacy penalty. Finally, an efficient algorithm is proposed to find optimal privacy-aware path in Hippocratic databases. Our work is grounded on modeling and analysis of purposes for Hippocratic databases and proposes enhancements to Hippocratic database systems in order to deal with inter-organizational business processes.

The remainder of the paper is structured as follows. Section 2 introduces the motivation of this paper based on a running example. Section 3 presents some background information on Hippocratic database systems. We introduce purpose directed graph with delegation in Section 4 and discuss how to characterize the privacy penalty and efficiently find the optimal solution in Section 5. We provide a brief survey of related work in Section 6. We conclude the paper in Section 7.

2 Motivation

We consider the following example throughout the paper.

Example: *Ebay is an online seller in Australia and provides an online catalogue to its customers who can search for the items they wish to buy. Once customers have decided to buy goods, Ebay needs to obtain certain personal information from customers to perform purchase transactions. This information includes name, shipping address, and credit card number. Ebay views purchase, its ultimate purpose, as a three-step process: credit assessment, delivery, and notification. Credit assessment relies on Credit Card Company (CCC). Delivery can be done either by a delivery company or the post office, while notification can be done by email or by mobile phone.*

Obviously, Ebay provides many ways to achieve the purchase service and each different method could require different data. An important principle is that enterprises should disclose to customers which data is collected and for what purpose. Also, enterprises should maintain minimal personal information necessary to fulfill the purpose for which the information was collected.

From the customers' point of view, they do not want to disclose more data than needed to get the desired service; rather, they want the process that best protects their privacy based on their preferences. Depending on the method of

notification, Ebay needs either an email address or a mobile phone number. For example, Jimmy, a professor plagued by spam, may treasure his email address and give away his business mobile phone number. Bob, a doctor whose mobile phone is always ringing, may have the opposite preference. Therefore, it is up to customers to decide how to get a service fulfilled on the basis of their personal feeling of any service customization.

Furthermore, if we consider the delivery service, Ebay could not fulfill the service by itself, but rather relies on a delivery company or post office. That means Ebay may outsource a large part of data processing to third parties participating in a single business process. However, the more the data is used, the more likely it might be disclosed, since the personal information is transmitted from one to another. This requires that enterprises maintain minimal personal information necessary to fulfill the purpose. Moreover, the partners chosen by Ebay might also be trusted differently by its potential customers. The burden of choice is on the human who must decide what to do on the basis of his/her personal feeling of trust of the enterprises. For instance, Albert may prefer to delivery by a delivery company, since it is fast; whereas, Bob may chose delivery by post office because it is safe. Different partners (sub-contractors) chosen for the same purpose may be with different trust levels. The choice of service customization has significant impact on the privacy of individual customers.

If we consider these factors, both the privacy cost and customer's trust should be considered as important factors in privacy security system when enterprises publish comprehensive privacy policies involving hierarchies of purposes, possibly spanning multiple partners. Formally, it can be stated as follows:

Minimal privacy cost: Is there a way to fulfill the purpose with minimal privacy cost?

Maximal customer's trust: Is there a way to fulfill the purpose with maximal trust between enterprises and customers?

Classical privacy-aware database systems such as Hippocratic databases do not consider these issues, we are interested in solutions that support customers and companies alike, so that companies can publish comprehensive privacy policies involving multiple service methods, possibly delegation of tasks and authorizations. Moreover, the solutions will allow customers to personalize services based on their own privacy sensitivities and their trust of partners who might contribute to the requested service.

3 Overview of Hippocratic databases

Hippocratic databases use *purpose* as a central concept [1]. A purpose describes the reason(s) for data collection and data access, which is stored in the database as a "special" attribute occurring in every table of the database. This attribute specifies the purpose (reason/goal) for which a piece of information can be used.

For example, Table 1 shows the schema of two tables, customer and order, that store the personal information including purposes. In particular, table

table	attribute
customer	purpose, customer-id, name, address, email, fax-number, credit-card-info
order	purpose, customer-id, transaction, book-info, status

Table 1. Database schema

table	attribute
privacy-polices	purpose, table, attribute, {external-receipts}, {retention-period}
privacy-authorizations	purpose, table, attribute, {authorized-users}

Table 2. Privacy metadata schema

customer stores personal information about customers, and table order stores information about the transactions between enterprises and their customers. Then, for each purpose and data item stored in the database, we have:

External-recipients: the actors to whom the data item is disclosed;

Retention-period: the period during which the data item should be maintained;

Authorized-users: the users entitled to access the data item.

Purpose, external recipients, authorized users, and retention period are stored in the database with respect to the metadata schema defined in Table 2. Specifically, the above information is split into separate tables: external-recipients and retention period are in the *privacy-policies table*, while authorized-users in the *privacy-authorizations table*. The purpose is stored in both of them. The privacy-policies table contains the privacy policies of the enterprise, while privacy-authorizations table contains the access control policies that implement the privacy policy and represents the actual disclosure of information. In particular, privacy-authorizations tables are derived from privacy-policies tables by instantiating each external recipient with the corresponding users. Therefore, Hippocratic database systems define one privacy-authorizations table for each privacy-policies table, and these tables represent what information is actually disclosed.

Hippocratic database system is an elegant and simple solution but does not allow for dynamic situations that could arise with web services and business process softwares. In such settings, enterprises may provide services in many different ways and may delegate the execution of parts of the service to third parties. This is indeed the case of a virtual organization based on business process for web service where different partners explicitly integrate their efforts into one process [15].

4 Purpose directed graph with delegation

Agrawal et al.[1] proposed a structure to split a purpose into multiple purposes and then stored them in the database. Karjoth et al.[16] used a directory-like notation to represent purpose hierarchies, which loses the logic relation between

a purpose and its sub-purposes. In particular, this notation does not distinguish if a sub-purpose is derived by AND or OR decomposition [19]. Assuming a purpose p is AND-decomposed into sub-purposes $p_1, \ldots, p_n$, then all of the sub-purposes must be satisfied in order to satisfy p. For example, Ebay AND-decomposes purchase into delivery, credit assessment, and notification, then all of the three sub-purposes have to be fulfilled for fulfilling purchase purpose. However, if a purpose p is OR-decomposed into sub-purposes $p_1, \ldots, p_n$, then one of the sub-purposes must be satisfied in order to satisfy p. For instance, Ebay further OR-decomposes delivery into direct delivery relying on delivery companies and delivery by post office (shown in Fig.1). In this way, only one of them could be necessary to fulfill the delivery purpose. In essence, AND-decomposition is used to define the process for achieving a purpose, while OR-decomposition defines alternatives for achieving a purpose.

Our approach is based on traditional goal analysis [18], and consists of decomposing purposes into sub-purposes through an AND/OR refinement. The idea is to represent purpose hierarchies with directed graphs.

Definition 1. *A purpose directed graph PDG is a pair (P, A), where P is a set of purposes and A is the set of arcs, each arc represents a hierarchical relation between the purposes.*

A purpose directed graph (PDG) can be used to represent goal models in goal-oriented requirements engineering approaches [7]. For our purposes, they represent the entire set of alternative ways for delivering a service required by customers. Such representations can also be used to model the delegations of tasks and authorizations in the security modeling methodology proposed by Giorgini et al.[14].

An enterprise could provide different methods to achieve a service or rely on different partners to achieve the same part of the service. In particular, Ebay relies on a delivery company, Worldwide Express (WWEx), for shipping books. Ebay needs to delegate customer's information, such as name and shipping address, to WWEx. In turn, WWEx depends on local delivery companies for door-to-door delivery. To this end, WWEx delegates customer information to the local delivery companies LDC_1, $\cdots$, LDC_n for door-to-door delivery. Consequently, different processes can be used to fulfill the required service. To capture this insight, we introduce the notion of path.

Definition 2. *Let $PDG = (P, A)$ be a purpose directed graph. A path from v_0 to v_m is defined as a sequence $W = (v_0, a_1, v_1, ..., a_m, v_m)$, where a_i is an arc from v_{i-1} to v_i for $i = 1, ..., m$.*

A purpose directed graph PDG is rooted if it contains a vertex v, such that all the vertices of PDG are reachable from v through a directed path. The vertex v is called a root of PDG.

For example, consider the purpose directed graph depicted in Fig.1. Each vertex is composed by two parts: a purpose identifier and an enterprise needed to fulfill the purpose, and each of the purposes represents the policies of a single

enterprise. The vertex 'purchase' is the root of the graph and purchase is the root-level purpose. Essentially, if a path $W = (v_0, a_1, v_1, ..., a_m, v_m)$ satisfies that v_0 is the root purpose and there exists no downward paths from v_m, we say the path is an essential path. An essential path represents a possible process through which an enterprise can fulfill the root purpose.

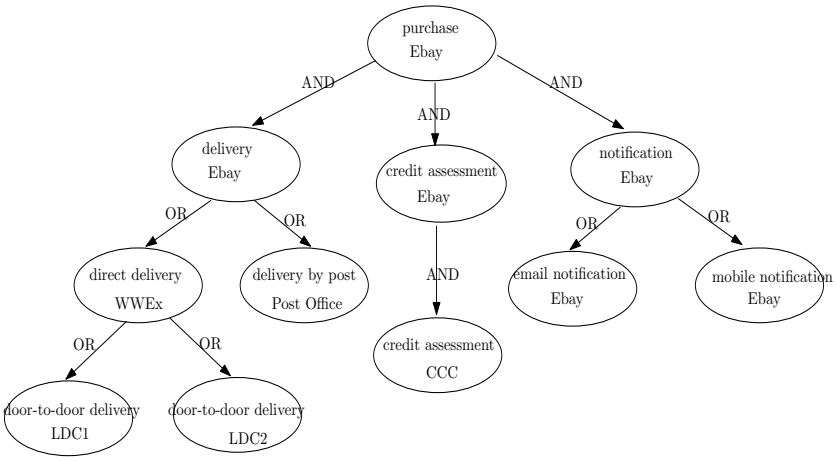

Fig. 1. Purpose directed graph

The enterprise-wide privacy policies are derived by looking at the Hippocratic database of each partner involved in the business process and merging them into a single purpose. Therefore, purposes can be recognized as the outcome of a process of refinements of goals in security requirements modeling methodologies[13]. The task delegation is indeed the case of a virtual organization based on business process for web service where different partners explicitly integrate their efforts into one process.

5 Finding Optimal Privacy-aware Path

Our goal is to decide which is the essential optimal privacy-aware path to fulfill the root purpose with respect to the customer's preference. This can be performed through the following quantitative analysis.

5.1 Objective characterization

Since our reference business model is that of virtual organizations, we assume that there will often be more than one way to deliver a service. Yet, they may differ in an important aspect, notably they may require different private data items, which incurs different privacy cost. Further, depending on each customer's

individual preferences, the same decomposition path might have significantly different trust values for different customers. In order to support quantitative analysis, we need to introduce the notion of privacy penalty.

Definition 3. *The privacy penalty of an arc a is defined as a pair $w_a = (\alpha, \beta)$, where α is the privacy cost and β is the customer's trust value on the arc a.*

Choice of α, β: The privacy penalty pair (α, β) on each arc can be pre-defined by asking the enterprises and customers to specify the level of privacy cost and trust they feel about the sub-suppliers. Since the personal information is transmitted from one to another, this may increase the danger of the leakage of personal information. Therefore, we use α to depict the privacy cost. Generally, we assume that there are different trust values based on the customer's personal feeling of the trust on different service custormizations. For example, Bob prefers mobile notification more than email notification because of the personal experience, so there is a high trust value on mobile notification.

Intuitively, the privacy penalty of a path should consist of two parts: one is the sum of the privacy cost on each arc and the other is the minimum trust among these arcs.

Definition 4. *Let $\mathcal{P} = (v_0, a_1, v_1, \cdots, a_m, v_m)$ be a path in the PDG. Then, the privacy penalty of the path $\omega_{\mathcal{P}} = \omega_{a_1} + \ldots + \omega_{a_m} = (\sum_{i=1}^{m}(\alpha_i), min_{i=1}^{m}(\beta_i))$, where $\omega_{a_i} = (\alpha_i, \beta_i)$, $i = 1, ..., m$.*

Essentially, a path represents a possible process through which an enterprise can fulfill a root purpose. For our purpose, we use the *sum* of the private cost of each arcs because we argue that the more a piece of data is used, the more likely it might be misused. The smaller the sum is, the less the privacy cost is. Therefore, *sum* measures are the ones that capture best one's intuitions on the cost of privacy. We also use the minimization function on trust values to get the smallest trust value on this path. The larger the value is, the more the trust is on this path. Our goal is to decide which is the process with the optimal privacy penalty (i.e., the minimal privacy cost and maximal trust value) to fulfill the root purpose with respect to the user's preferences. In order to describe the user's preference, we next introduce a flexible objective function.

Flexible objective function: If the privacy penalty on the arc a is defined as $w_a = (\alpha, \beta)$, we introduce the following objective function to balance the privacy cost and customer trust with a preference coefficient γ $(0 \leq \gamma \leq 1)$.

$$alt(a) = \gamma \times \alpha + (1 - \gamma) \times \beta \tag{1}$$

The choice of parameter γ depends on the customer's preference. If the customer cares whether data are disclosed at all, then γ may be set with a value in the interval $0.5 \leq \gamma \leq 1$. On the other hand, if the customer stresses more on trust, then γ can be set with a value between 0 and 0.5.

In addition to the objective function, we propose to decompose purposes into sub-purposes through an AND/OR decomposition. In essence, AND-decomposition

is used to define the process for achieving a purpose, while OR-decomposition defines alternatives for achieving a purpose. Normally, the node purpose can be either AND-decomposed or OR-decomposed. A decomposition arc is either an OR-arc or an AND-arc.

Definition 5. *Let $PDG = (P, A)$ be a purpose directed graph, for each vertex $v \in P$, we denote $OUT(v) = OUT_{or}(v) \cup OUT_{and}(v)$ as the set of all successors of v, where $OUT_{or}(v)$ refers to all successors connecting v with OR-arcs, and $OUT_{and}(v)$ stores all successors connecting v with AND-arcs. Especially, if $OUT(v) = \emptyset$, we say the vertex v is a leaf of PDG.*

For example, in Fig. 2 the root purpose r is AND-decomposed into three sub-purposes: delivery, credit assessment and notification, then $OUT(r) = OUT_{and}(r)$ = {delivery, credit assessment and notification}. Further, considering the node v with purpose 'mobile notification', since $OUT(v) = \emptyset$, then the node 'mobile notification' is a leaf of the purpose directed graph.

5.2 The algorithm

In this section, we present efficient algorithms to track the optimal path that the enterprises need to fulfill a purpose. Next, we analyze two situations in finding the optimal privacy-aware path.

Case 1: if the root purpose is OR-decomposed, the algorithm consists of following steps:

1. To contract each vertex v with all its successors in $OUT_{and}(v)$ to a compound vertex v_c; suppose $OUT_{and}(v) = \{v_1, \cdots, v_k\}$, we define $cost[v_c] = \sum_{i=1}^{k} \alpha(v, v_i)$, $trust[v_c] = \min_{i=1}^{k} \beta(v, v_i)$;
2. To transfer the purpose directed graph PDG into $\overline{PDG}$ with no AND-arcs and find the optimal path $\overline{p}$ using function $optimal\_path(\overline{PDG})$;
3. If the optimal path of $\overline{PDG}$ contains a compound vertex (or vertices), then expand the compound vertex (or vertices) on $\overline{p}$ to become the optimal solution of PDG.

In Algorithm 1, $\alpha(u, v)$ represents the privacy cost between the two nodes u and v, and $\beta(u, v)$ refers to the trust value on the arc (u, v). For each leaf vertex, Sum function is used to track the distance between the leaf and the root, while $predecessor[]$ records all predecessor vertices of the leaf, and $previous[]$ records the vertices on the optimal path from the leaf to the root. alt on line 10 is the objective function with the preference coefficient γ. If $\gamma \geq 0.5$, it means customers prefer more on privacy protection, then the minimal objective value is needed depending on the minimization function; while if $\gamma < 0.5$, it means customers prefer more on trust, then the maximal objective value is needed depending on the maximization procedure.

Case 2: if the root purpose is AND-decomposed, in order to design efficient algorithms to determine the process by which a service can be delivered with optimal privacy penalties, we need the definition of sub-purpose directed graph.

Algorithm 1: *optimal_path*(PDG, OR_r)

Input: a purpose directed graph PDG with OR-decompoded root r.
Output: The optimal path D

1. Contract each vertex v with all its successors in $OUT_{and}(v)$ to the compound vertex v_c
2. Transfer PDG into $\overline{PDG}$
3. $\overline{p} = optimal\_path(\overline{PDG})$
4. If $\overline{p}$ contains compound vertex(vertices),
5. expand the compound vertex(vertices) on $\overline{p}$ to p,
6. $D = p$
7. else
8. $D = \overline{p}$

function: *optimal_path*$(\overline{PDG})$:

Input: $\overline{PDG}$ with root purpose r and leaves $v_1, \ldots, v_k$, pre-defined privacy cost and trust function $\alpha(*,*)$, $\beta(*,*)$, and preference coefficient $0 \leq \gamma \leq 1$

0 for each vertex v (not a compound vertex) in $\overline{PDG}$:
1 $cost[v] := 0$
2 $trust[v] := \infty$
3 for each leaf v_i $(i = 1, \cdots, k)$
4 $Sum(v_i) := 0$, $previous[v_i] := \{v_i\}$
5 while $r \notin predecessor[v_i] = \{u_{i_1}, \cdots, u_{i_s}\}$
6 {
7 for each u_{i_j} $(1 \leq j \leq s)$
8 $cost(u_{i_j}, v_i) := cost[v_i] + cost[u_{i_j}] + \alpha(u_{i_j}, v_i)$
9 $trust(u_{i_j}, v_i) := min\{trust[v_i], trust[u_{i_j}], \beta(u_{i_j}, v_i)\}$
10 $alt(u_{i_j}, v_i) := \gamma \times cost(u_{i_j}, v_i) + (1 - \gamma) \times trust(u_{i_j}, v_i)$
11 if $\gamma \geq 0.5$ /* prefer cost */
12 let $alt(u_{i_m}, v_i) = \min_{j=1}^{s} alt(u_{i_j}, v_i)$
13 $previous[v_i] := previous[v_i] \cup \{u_{i_m}\}$
14 $Sum(v_i) := Sum(v_i) + alt(u_{i_m}, v_i)$
15 $v_i := u_{i_m}$
16 if $\gamma < 0.5$ /* prefer trust */
17 let $alt(u_{i_m}, v_i) = \max_{j=1}^{s} alt(u_{i_j}, v_i)$
28 $previous[v_i] := previous[v_i] \cup \{u_{i_m}\}$
29 $Sum(v_i) := Sum(v_i) + alt(u_{i_m}, v_i)$
20 $v_i := u_{i_m}$
21 }
22 /*end while and all paths from the leaf to the root are found*/
23 if $\gamma \geq 0.5$
24 assume $Sum(v_t) = min_{i=1}^{k} Sum(v_i), (1 \leq t \leq k)$
25 output $previous[v_t]$
26 if $\gamma < 0.5$
27 assume $Sum(v_t) = max_{i=1}^{k} Sum(v_i), (1 \leq t \leq k)$
28 output $previous[v_t]$
29 end function

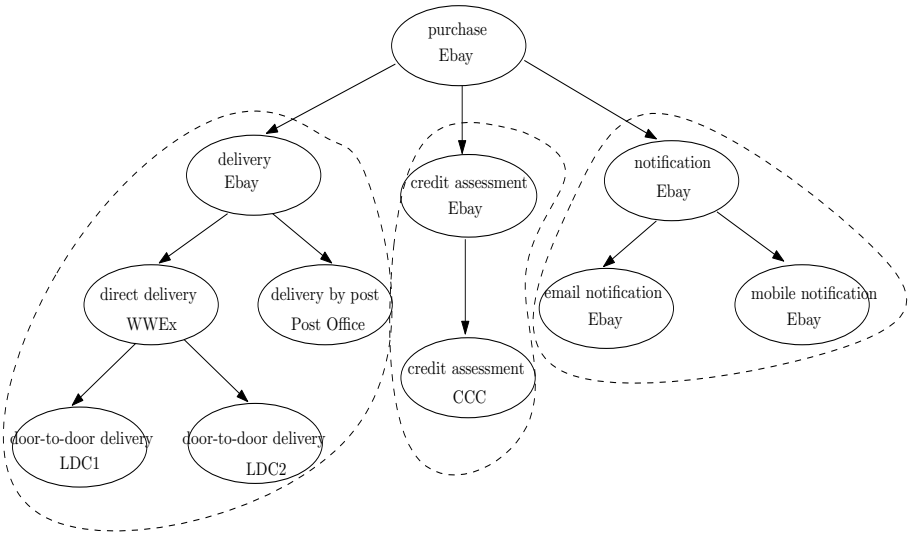

Fig. 2. *sub_PDG* in Purpose directed graph

Definition 6. *Let $PDG = (P, A)$ be a purpose directed graph, if the root purpose r is AND-decomposed into several sub-purposes, then each sub-purpose with all its descendants form a sub-purpose directed graph of PDG, and we denote it by sub_PDG. Essentially, if the root of the sub_PDG is further AND-decomposed into several sub-purposes, then each sub-purpose with all its descendants form a sub-purpose directed graph of sub_PDG, which is also a sub-sub-purpose directed graph of PDG, and we denote it by sub_(sub_PDG).*

For example, in Fig. 2 Ebay AND-decomposes purpose purchase into three sub-purposes: delivery, credit assessment and notification. According to the definition of sub-purpose directed graph, the purpose delivery with all its decedents consist of a sub-purpose directed graph. The same situation applies to the other two sub-purposes, so there are three sub-purpose directed graphs as in Fig. 2 (circled in broken line). Since in each sub-purpose directed graph, the root is further OR-decomposed, there is no sub-sub-purpose directed graph in Fig. 2.

For the sake of simplicity, we assume that the root of each *sub_(sub_PDG)* is OR-decomposed. In this case, the algorithm consists of following steps:

1. To decompose the purpose directed graph PDG into several sub-purpose directed graphs.
2. For each sub-purpose directed graph *sub_PDG* with root purpose r,
   (a) if the root purpose r is OR-decomposed, run algorithm *optimal_path*(*sub_PDG*, OR_r) to find the optimal path in *sub_PDG*;
   (b) if the root purpose r is AND-decomposed, further decompose the *sub_PDG* into several sub-sub-purpose directed graphs, then run algorithm *optimal_path* (*sub_(sub_PDG)*, OR_r') to find the optimal path in each *sub_(sub_PDG)*

with root r'. Combine all the optimal paths of each $sub\_(sub\_PDG)$ into the optimal solution of sub_PDG.

3. To combine all the optimal paths of each sub_PDG into the optimal solution of PDG.

Algorithm 2: $optimal\_path(PDG, AND\_r)$

Input: A purpose directed graph PDG with AND-decomposed root
Output: The optimal path D

```
1.  decompose PDG into several sub_PDG
2.  for each sub_PDG with root r
3.   if the root r is OR-decomposed in sub_PDG
4.     run algorithm optimal_path(sub_PDG, OR_r)
5.      output p_or = optimal_path(sub_PDG)
6.   if the root r is AND-decomposed in sub_PDG
7.     further decompose the sub_PDG into several sub_(sub_PDG)s
2.      for each sub_(sub_PDG) with root r'
8.       run algorithm optimal_path(sub_(sub_PDG), OR_r')
9.        output p_and = optimal_path(sub_PDG)
10.  D = (∪p_or) ∪ (∪p_and)
```

In Algorithm 2, p_{or} refers to the optimal path of sub_PDG with an OR-decomposed root, while p_{and} refers to the optimal path of sub_PDG with an AND-decomposed root.

Until here, ether the root purpose is AND-decomposed or OR-decomposed, we can find the optimal path under any specific value of γ through our algorithms. If $\gamma \geq 0.5$, it means the customer stresses more on privacy cost. The optimal solution to satisfy the customer's preference is the optimal path with the minimal objective value when varying the value of γ. Our method is to search all the possible optimal path based on the value of γ from 0.5 to 1 by the interval of 0.01. Then, the optimal path with the minimal objective value will be chosen as the optimal solution. For the situation $\gamma < 0.5$, we search all the possible optimal path based on the value of γ from 0 to 0.5 by the interval of 0.01. Then, the optimal path with the maximal objective value will be chosen as the optimal solution, since the customer prefers more on trust.

6 Related work

Our work is related to several topics in the area of privacy and security for data management, namely privacy policy specification, privacy-preserving data management systems and multilevel secure database systems. We now briefly survey the most relevant approaches in these areas and point out the differences of our work with respect to these approaches.

The W3Cs Platform for Privacy Preference (P3P) [23] allows web sites to encode their privacy practice, such as what information is collected, who can

access the data for what purposes, and how long the data will be stored by the sites, in a machine-readable format. P3P enabled browsers can read this privacy policy automatically and compare it to the consumers set of privacy preferences which are specified in a privacy preference language such as a P3P preference exchange language (APPEL) [10], also designed by the W3C. Even though P3P provides a standard means for enterprises to make privacy promises to their users, P3P does not provide any mechanism to ensure that these promises are consistent with the internal data processing. By contrast, the work in our paper provides an effective strategy to maximize privacy protection. Further, we allow customers to express their trust preferences associated with each partner of the business process in order to achieve maximal customer trust.

Byun et al. presented a comprehensive approach for privacy preserving access control based on the notion of purpose [9, 8]. In the model, purpose information associated with a given data element specifies the intended use of the data element, and the model allows multiple purposes to be associated with each data element. The granularity of data labeling is discussed in detail in [9], and a systematic approach to implement the notion of access purposes, using roles and role-attributes is presented in [8]. Similar to our approach, they introduce purpose hierarchies in order to reason on access control. Their hierarchies are based on the principles of generalization and specification and are not expressive enough to support complex strategies defined by enterprises. However, we organize purposes into purpose directed graph through AND/OR decomposition, which supports the delegation of tasks and authorizations when a host of partners participating in the business process provides different ways to achieve the same service. We also present an efficient method to automatically derive the optimal way of authorizations needed to achieve a service from enterprise privacy policies.

The concept of Hippocratic databases, incorporating privacy protection within relational database systems, was introduced by Agrawal et al.[1]. The proposed architecture uses privacy metadata, which consist of privacy policies and privacy authorizations stored in two tables. LeFevre et al.[2] enhance Hippocratic databases with mechanisms for enforcing queries to respect privacy policies stated by an enterprise and customer preferences. In essence, they propose to enforce the minimal disclosure principle by providing mechanisms to data owners that control as who can access their personal data and for which purpose. Although the work on the Hippocratic databases[1, 2] is closely related to ours, our approach has some notable differences. First, we introduce more sophisticated concepts of purpose, i.e., purposes are organized in purpose directed graph through AND/OR decomposition. The second difference is that Hippocratic databases does not allow to distinguish which particular method is used; whereas, we discuss the situations that could arise with web services and business process software. Third, we provide an efficient method to automatically derive the optimal way of authorizations needed to achieve a service from enterprise privacy policies.

7 Conclusions

In this paper, we analyze the purposes behind the design of Hippocratic database systems, and organize them in hierarchal manner through AND/OR decomposition. We apply the purpose directed graph to characterize the ways the enterprised need to achieve a service which may rely on many different partners. Specially, the selection of the partners and the identification of a particular plan to fulfill a purpose is driven by the customer's preference. We use a goal-oriented approach to analyze privacy policies of the enterprises involved in a business process, in which one can determine the minimum disclosure of data for fulfilling the root purpose with respect to customer's maximum trust. On the basis of the purpose directed graph derived through a goal refinement process, we provide efficient algorithms to determine the optimal privacy-aware path for achieving a service. This allows to automatically derive access control policies for an inter-organizational business process from the collection of privacy policies associated with different participating enterprises.

References

1. R. Agrawal, J. Kiernan, R. Srikant and Y. Xu, Hippocratic Databases. In: Proceedings of VLDB'02, pp. 143-154. Morgan Kaufmann, San Francisco (2002)
2. R. Agrawal, A. Evfimievski and R. Srikant, Information sharing across private databases. In: Proceedings of SIGMOD'03, pp. 86-97. ACM Press, New York (2003)
3. P. Ashley, C.S. Powers and M. Schunter, Privacy promises, access control, and privacy management. In: Third International Symposium on Electronic Commerce (2002)
4. M. Backes, B. Pfitzmann and M. Schunter, A Toolkit For Managing Enterprise Privacy Policies. In: Proceedings Of Esorics'03, Lncs 2808, pp. 162-180. Springer, Berlin Heidelberg New York (2003)
5. E. Bertino, E. Ferrari and A.C. Squicciarini, Trust-X: A Peer-to-Peer Framework for Trust Establishment. IEEE Trans. Knowl. Data Eng. 16(7): 827-842 (2004)
6. M. Blaze, J. Feigenbaum, and J. Lacy, Decentralized trust management, in Proc. IEEE Symp. Security Privacy, 1996, pp. 164-173.
7. P. Bresciani, P. Giorgini, F. Giunchiglia, J. Mylopoulos, and A. Perini, TROPOS: An agent-oriented software development methodology. JAAMAS 8(3), 203-236 (2004)
8. J.W. Byun, E. Bertino, and N. Li: Purpose based access control of complex data for privacy protection. In: Proceedings of SACMAT'05, pp. 102-110. ACM Press, New York (2005)
9. J.W. Byun, E. Bertino and N. Li, Purpose based access control for privacy protection in relational database systems. *Technical Report* 2004-52, Purdue University.
10. L. Cranor, M. Langheinrich, M. Marchiori, and J. Reagle: The platform for privacy preferences 1.0 (P3P1.0) specification. W3C recommendation (2002). http://www.w3.org/TR/P3P/
11. E. Ferrari and B.M. Thuraisingham, Security and privacy for web databases and services. In: Proceedings of the 9th International Conference on Extending Database Technology, LNCS 2992, pp. 17-28. Springer, New York (2004).

12. T. Finin and A. Joshi, Agents, trust, and information access on the semantic web, ACM SIGMODRec. , vol. 31, no. 4, pp. 30-35, Dec. 2002.
13. P. Giorgini, F. Massacci, J. Mylopoulos and N. Zannone, Requirements Engineering meets Trust Management: Model, Methodology, and Reasoning. In Proc. of iTrust'04, LNCS 2995, pp. 176-190. Springer-Verlag, 2004.
14. P. Giorgini, F. Massacci, J. Mylopoulos and N. Zannone: Modeling security requirements through ownership, permission and delegation. In: Proceedings of RE'05, pp. 167-176. IEEE Press, Lausanne (2005)
15. C. Handy: Trust and the virtual organization. Harv.Bus.Rev.73,40-50(1995)
16. G. Karjoth, M. Schunter and M. Waidner: Platform for enterprise privacy practices: privacy-enabled management of customer data. In: Proceedings of PET'02, LNCS 2482, pp. 69-84. Springer, Berlin Heidelberg New York (2002)
17. F. Massacci and N. Zannone: Privacy is linking permission to purpose. In: Proceedings of the 12th InternationalWorkshop on Sec. protocols (2004)
18. N.J., Nilsson, Problem solving methods in AI. McGraw-Hill, 1971.
19. N.J., Nilsson, Principles of Artificial Intelligence, Morgan Kaufman, 1994.
20. K.E. Seamons, M. Winslett, T. Yu, L.Yu and R. Jarvis: Protecting privacy during on-line trust negotiation. In: Proceedings of PET'02, LNCS 2482, pp. 129-143. Springer, Berlin Heidelberg New York (2002)
21. A. Tumer, A. Dogac and H. Toroslu: A Semantic based Privacy framework forweb services. In: Proceedings of ESSW'03 (2003)
22. M. Yasuda, T. Tachikawa and M.Takizawa: Information flow in a purpose-oriented access control model. In: Proceedings of ICPADS'97, pp. 244-249. IEEE Press, Lausanne (1997)
23. World Wide Web Consortium (W3C). A P3P Preference Exchange Language 1.0 (APPEL 1.0). Available at www.w3.org/TR/P3P-preferences.